



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2016

Drei Thesen zu Visualisierungspraktiken in den Digital Humanities

Bubenhofer, Noah

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-126501>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution-NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0) License.

Originally published at:

Bubenhofer, Noah (2016). Drei Thesen zu Visualisierungspraktiken in den Digital Humanities. *Rechtsgeschichte*, 24:351-355.

Rechtsgeschichte Legal History

www.rg.mpg.de

<http://www.rg-rechtsgeschichte.de/rg24>

Rg **24** 2016 351 – 355

Noah Bubenhofer

Drei Thesen zu Visualisierungspraktiken in den Digital Humanities

Dieser Beitrag steht unter einer
Creative Commons cc-by-nc-nd 3.0



Noah Bubenhofer

Drei Thesen zu Visualisierungspraktiken in den Digital Humanities

In den Wissenschaften dienen Visualisierungen dazu, 1. eine Idee in eine grafische Form zu überführen, um sie verständlicher zu machen, 2. statistische Ergebnisse zu präsentieren oder 3. Daten zu explorieren. In den Digital Humanities wird vor allem der dritten Funktion großes Potenzial zugeschrieben: Mit visuellen Analysemethoden sollen große, komplexe und manchmal unstrukturierte Daten so hergerichtet werden, dass sie aus Distanz »gelesen« werden können (»Distant Reading«, vgl. MORETTI 2009). So werden z. B. Zusammenhänge zwischen Akteuren oder anderen Entitäten als Netzwerk modelliert und visualisiert oder Informationen mit geografischem Bezug kartiert. Diese Visualisierungen, auch »Diagramme« genannt, bestehen im besten Fall aus folgenden grundsätzlichen Arbeitsschritten:

- Für Elemente in den Daten, die von Interesse sind, wird ein Set von grafischen Zeichen geschaffen, die in *ikonischen* Verhältnissen zum Gemeinten stehen – zusammen mit Regeln für die Transformation der Elemente in die grafischen Zeichen (Beispiel: Knoten stehen für Akteure, Kanten stehen für Beziehungen).
- Die Anwendung der definierten Regeln auf Daten führt zu einem Diagramm (Beispiel: Netz von Knoten und Kanten). Entscheidend für die Visualisierung digitaler Daten ist hier, dass diese Regeln programmiertechnisch als *Algorithmus* implementiert werden können. Damit können grundsätzlich beliebige, ähnlich strukturierte Daten, in jedem Umfang, visualisiert werden.
- Nun kann mit dem Diagramm »operiert« werden, wobei im besten Fall neue Erkenntnisse über die Daten gewonnen werden, da das Diagramm Eigenschaften oder Zusammenhänge sichtbar macht, die vorher nicht sichtbar waren (Beispiel: Knoten im Netz, die besonders viele Verbindungen zu anderen Knoten aufweisen und deshalb als zentrale Knoten interpretiert werden können).

Diese »diagrammatischen« Aspekte von Visualisierungen wurden und werden aus semiotischer Perspektive in ihrer ganzen Komplexität beleuchtet (PEIRCE 1994; KRÄMER 2009; BREDEKAMP/SCHNEIDER/DÜNKEL 2008 und viele mehr). Auf der anderen Seite werden die Herausforderungen

der Konzeption, Implementierung und Evaluation von visuellen Analysesystemen vor allem in neuerer Zeit hauptsächlich unter dem Label »Visual Analytics« behandelt (CHEN/HÄRDLE/UNWIN 2008; KEIM u. a. 2010; THOMAS/COOK 2005). Daraus ergab sich in den letzten Jahren eine deutliche Professionalisierung in Form von wissenschaftlichen Disziplinen, die sich diesen konzeptionellen und vor allem technisch herausfordernden Aufgaben stellen.

Für die digitalen Geisteswissenschaften ist das eigentlich eine erfreuliche Entwicklung: Es stehen Tools und Algorithmen zur Verfügung, um große Datenmengen bezüglich bestimmter Aspekte zu visualisieren und damit interpretierbar zu machen.

Gleichzeitig wirft die Anwendung von visuellen Analysemethoden in den Geisteswissenschaften eine Reihe von kritischen Fragen auf, die kaum diskutiert werden und die ich in Form von drei – provokativen – Thesen antippen möchte:

1. Visuelle Analysemethoden folgen utilitaristischen und positivistischen wissenschaftlichen Denkstilen, die nur bedingt mit den Geisteswissenschaften kompatibel sind.
2. Die Geisteswissenschaften unterschätzen die Bedeutung der programmiertechnischen und algorithmischen Grundlagen der verwendeten visuellen Analysemethoden.
3. Visuelle Analysetools für die Digital Humanities laufen Gefahr, traditionellen Interpretationsmethoden nachzueifern statt die Chancen des datengeleiteten Paradigmas zu nutzen.

Erstens: Mit »Illuminating the Path« gaben 2005 THOMAS und COOK die Forschungsagenda für »Visual Analytics« heraus, finanziert durch das US Department of Homeland Security – eine Reaktion auf die Terroranschläge 9/11 in den USA. Das Ziel: »the development of advanced information technologies to support the homeland security mission« (ebd.: i). Visuelle Analysemethoden sollen also dabei behilflich sein, im Datenwust der digitalen Welt Spuren zu finden, die auf kriminelle/terroristische Aktivitäten hindeuten: »identifying a plot or threat that is hinted at, but not clearly communicated, by a small subset of the documents in the collection« (GÖRG/LIU/STASKO

2014). Auch wenn in der Folge visuelle Analysemethoden in völlig unterschiedlichen Gebieten eingesetzt worden sind, ist doch immer das Leitmotiv deutlich hörbar. Visual Analytics zielt auf die Entwicklung von »effective analysis tools« (KEIM u. a. 2010: 2), »to turn information overload [...] into a useful asset« (KEIM u. a. 2006), indem sie die Wahrheit in den Daten ans Licht holen. Systeme, die dieses Ziel verfolgen, können – und müssen in logischer Konsequenz – anschließend darauf hin evaluiert werden, wie gut sie dieses Ziel erfüllen. Dafür testet man die Methode anhand eines Testdatensatzes, in dem eine »ground truth« verborgen ist, die gefunden werden soll (KEIM u. a. 2010: 139; GÖRG/LIU/STASKO 2014: 344).

Trotz aller Vielfalt von Analyse- und Interpretationsmethoden in den Geisteswissenschaften mutet dieses Verständnis von Datenanalyse grundsätzlich etwas befremdlich an. Seien es Paradigmen wie Poststrukturalismus, Dekonstruktivismus, Diskursanalyse, Kulturanalyse, Gender Studies, Systemtheorie usw., in Daten, Quellen oder Artefakten wird normalerweise keine »ground truth« angenommen, die es bloß zu finden gilt. In der Linguistik werden beispielsweise Textkorpora und entsprechende Analysemethoden, darunter auch visuelle Analysemethoden, mit Gewinn benutzt, um Diskurse zu untersuchen und sprachliche Muster mit sozialem und kulturellem Handeln in Verbindung zu bringen (BUBENHOFER 2009; SPITZMÜLLER/WARKE 2011; ANGERMULLER 2014; FELDER/MÜLLER/VOGEL 2011; SCHARLOTH/EUGSTER/BUBENHOFER 2013; MAUTNER 2012). Explorative, datengeleitete Methoden helfen dabei, in den Daten Zusammenhänge zu entdecken, die vor dem Hintergrund einer Theorie gedeutet werden. Die theoretische Basis, die Methodologie und deren Operationalisierung und Implementierung als Analyseinstrument bilden zusammen mit der Forscherin/dem Forscher und dem disziplinären Setting ein Gewebe, das die Deutung formt. Ob wir im Rauschen eine »wissenschaftliche Tatsache« (Ludwig Fleck) erkennen, hängt vom Denkkollektiv ab, in dem wir uns bewegen: »Alles Erkennen ist ein Prozeß zwischen dem Individuum, seinem Denkstil, der aus der Zugehörigkeit zu einer sozialen Gruppe folgt, und dem Objekt« (FLECK 1983: 168). Die Evaluationsmethoden, die innerhalb einer wissenschaftlichen Disziplin entwickelt werden, dienen demnach genau dazu: zu entscheiden, ob innerhalb dieser Disziplin – und nur dieser Disziplin – ein Instrument in der Lage ist, im

Rauschen der Daten das zu entdecken, was nach Ansicht des Denkkollektivs eine wissenschaftliche Tatsache ist. Die Instrumente der Visual Analytics werden – innerhalb der Geisteswissenschaften angewendet – dabei jedoch zu anderen Instrumenten, die den Regeln geisteswissenschaftlicher Denkkollektive folgen. Ich zweifle, ob sie da, einem utilitaristischen Topos folgend (GEOFFEY 2014: 21; BÖHME 2006: 55), effektive Analysetools sind, die den »Informationsüberfluss« zähmen können. Nach geisteswissenschaftlicher Logik sind sie vielleicht eher dann »nützlich«, wenn sie reiche Nahrung für die Deutung bieten, etwa im Sinne einer »Dichten Beschreibung« (GEERTZ 1987).

Zweitens: Zur Verarbeitung und Analyse von digitalen Daten kommen Algorithmen, statistische Maße, Berechnungen etc. zur Anwendung. Hinter einer Netzwerkanalyse beispielsweise stecken eine Vielzahl von Schritten der Datenaufbereitung und ein Algorithmus, mit dem die Knoten und Kanten des Netzes definiert werden. Schließlich ist es ein weiterer Algorithmus, der das Netz visualisiert und dabei z.B. physikalische Kräfte in Form von Sprungfedern zwischen den Knoten simuliert, um die Knoten auf einer Fläche anzuordnen (PFEFFER 2010).

Bei jedem der Schritte müssen Entscheidungen getroffen werden, die sich auf alle weiteren Schritte auswirken – und selbstverständlich auch auf die Interpretation der Daten. Es ist also gefährlich, solche Analyseinstrumente als Black Boxes zu nutzen, als Nutzer oder Nutzerin also nicht Bescheid über die Funktionsweise des Instruments zu wissen. Dies ist im Fokus der Digital Humanities, wenn auch das Problem längst nicht gelöst ist, da damit die Anforderungen an die methodischen und theoretischen Kompetenzen der Forscherin/des Forschers stark steigen (BERRY 2012: 75; RIEDER/RÖHLE 2012; BUBENHOFER/SCHARLOTH 2015).

Relativ wenig Gewicht wurde bislang aber der programmiertechnischen Implementierung dieser Algorithmen beigemessen: Macht es einen Unterschied, ob der Algorithmus zur Netzdarstellung in der Programmiersprache C oder in Javascript umgesetzt wird?

Ich möchte darauf aufmerksam machen, dass Programmieren eine zutiefst kulturell geprägte Tätigkeit ist (COLEMAN 2012; MANOVICH 2014; FORD 2015). Mit der Wahl einer Programmiersprache und einer Programmierumgebung schreibt man sich in eine Programmierkultur ein,

die von bestimmten Idealen und Topoi geprägt ist. Deutlich wird das beispielsweise, wenn man sich den Wandel der Programmierkulturen betrachtet, die etwa mit der OpenSource-Bewegung und kollaborativen Webplattformen (»Web 2.0«) einher geht. Im Bereich der Datenvisualisierung sind heute Javascript-Bibliotheken wie »D3« (BOSTOCK/OGIEVETSKY/HEER 2011) prägend: Damit kann auf Basis von bereits bestehendem Code, der offen ist und im Web distribuiert wird, leicht eine neue Visualisierungsanwendung programmiert werden, die ohne komplizierte Anpassungen sofort auf jedem Rechner mit jedem Betriebssystem im Webbrowser funktioniert. Noch vor wenigen Jahren war die Programmierung solcher Tools weit aufwändiger und führte zu Anwendungen, deren Code nicht (ohne weiteres) einsehbar war und die differenzierte Systemvoraussetzungen verlangten, um ausgeführt werden zu können. Damit ist die technische Hürde, sich selber ein Analysetool zu programmieren, deutlich niedriger geworden, was ein gewichtiger Vorteil ist, um der Forderung nach White Boxes nachkommen zu können. Gleichzeitig ist aber auch klar, dass ein deutlicher Zusammenhang zwischen Programmierkulturen und darin entstehenden Programmen besteht, der sich auf die Analysearbeit mit diesen Programmen auswirkt, etwa indem ihnen dezentrale, unkoordinierte Remix-Arbeitsweisen inhärent sind (vgl. dazu auch BERRY 2008). Es ist für die Digital-Humanities-Forscherin relevant, ob sie mit einem Tool arbeitet, das in C oder Javascript erstellt worden ist.

Drittens: In den Visual Analytics gibt es ein »Mantra«, das ein »gutes« Werkzeug beschreibt: »Overview first, zoom and filter, then details-on-demand« (SHNEIDERMAN 1996). Die Visualisierung soll also sowohl den Überblick über die Daten als auch den Fokus auf das einzelne Datum erlauben. Bei der Arbeit mit Textdaten wird dies oft so interpretiert, dass mit dem Analyseinstrument letztlich jederzeit auf das einzelne Textdokument zugegriffen werden können muss.

Diese Forderung scheint auf den ersten Blick plausibel und hat gewiss auch ihre Berechtigung. Nur besteht die Gefahr, dabei gerade den Vorteil quantitativer, datengeleiteter Analysen von Daten aus dem Blick zu verlieren. Franco Moretti forderte aus literaturwissenschaftlicher Sicht: »what we really need is a little pact with the devil: we know how to read texts, now let's learn how *not* to read them« (MORETTI 2000: 57). Die Digital Human-

ities, genauer vielleicht die maschinellen Textanalysemethoden, bringen den Vorteil, die Einheit des Textes aufzubrechen (BUBENHOFER 2009: 105). Ziel ist es, durch eine Aufbereitung des Textkorpus, ggf. durch eine Visualisierung bestimmter Strukturen, ein emergentes Phänomen sichtbar zu machen. Ein Phänomen, das sich also nicht aus der Summe der Einzelbeobachtungen erklären lässt, sondern mehr als das ist. Ein mit statistischen Methoden erstelltes Kollokationsprofil eines Wortes ist ein solches Phänomen: Es fasst das statistisch überzufällige Kovorkommen von Wörtern in großen Textkorpora so zusammen, dass daraus ein Bild entsteht, wie das Wort in den Daten normalerweise verwendet wird (EVERT 2009). In weiteren Abstraktionsschritten können die Kollokationsprofile maschinell nach Ähnlichkeit gruppiert werden, um semantische Felder (KEIBEL/BELICA 2007) zu berechnen oder Lexempaare zu finden, die in gleichen semantischen Relationen zueinander stehen (etwa nach MIKOLOV u. a. 2013). Der Einzelbeleg ist dabei unbedeutend – er kann dem emergenten Profil sogar widersprechen. Interessant ist die zusammenfassende Darstellung, die genug ungenau ist, um das Wesentliche vom Unwesentlichen zu trennen.

Wenn der Forscher oder die Forscherin bei der Analyse der Daten doch meint, ein Blick in die Einzeltexte sei notwendig, um die quantitativ aufbereiteten Daten deuten zu können, dann ist die quantitative Aufbereitung noch keine genügend gute Operationalisierung des zu untersuchenden Phänomens. Es mag z. B. sein, dass sich im Kollokationsprofil Daten finden, die wir nicht als generellen Aspekt der Semantik des entsprechenden Wortes akzeptieren würden – beispielsweise weil sie auf einen häufig genannten Buch- oder Filmtitel in den Daten zurück gehen, was sichtbar wird, wenn man sich die Einzelbelege ansieht. Dies bedeutet jedoch, dass die Operationalisierung von Semantik eines Wortes mittels Kollokationsprofilen noch nicht ideal ist, sei es, dass die Berechnung noch zu wenig gut modelliert ist oder aber, dass die Daten zu wenig ausgewogen oder umfangreich sind. Wenn diese Probleme jedoch gelöst sind, müsste der Blick in die Einzelbelege nicht mehr notwendig sein.

Wenn trotz dieser Überlegungen der Wunsch bleibt, immer wieder einen Blick in die Details zu werfen, verbirgt sich dahinter vielleicht ein wissenschaftlicher Gestus, der nach wie vor dem Einzeltext huldigt und in der quantitativen Datenanalyse

ein Mittel sieht, diese Aufgabe, das Lesen von Einzeltexten, schneller erledigen zu können. Dies können quantitative Methoden generell und visuelle Analysemethoden im Besonderen nicht leis-

ten. Diese Methoden ermöglichen neue Fragestellungen, nicht Lösungen für die alten Fragen.



Bibliographie

- ANGERMÜLLER, JOHANNES (2014), ›Der‹ oder ›das‹ Korpus? Perspektiven aus der Sozialforschung, in: Diskursforschung Ein interdisziplinäres Handbuch, Berlin, Boston: De Gruyter
- BERRY, DAVID (2008), Copy, Rip, Burn: The Politics of Copyleft and Open Source, London: Pluto Press
- BERRY, DAVID M. (Hg.) (2012), Understanding Digital Humanities, Basingstoke: Palgrave Macmillan
- BÖHME, GERNOT (2006), Technical Gadgetry: Technological Development in the Aesthetic Economy, in: Thesis Eleven 86/1, 54–66
- BOSTOCK, MICHAEL, VADIM OGIEVETSKY, JEFFREY HEER (2011), D3: Data-Driven Documents, IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis), <http://vis.stanford.edu/papers/d3>
- BREDEKAMP, HORST, BIRGIT SCHNEIDER, VERA DÜNKEL (Hg.) (2008), Das Technische Bild: Kompendium zu einer Stilgeschichte wissenschaftlicher Bilder, Berlin: Akademie Verlag
- BUBENHOFER, NOAH (2009), Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse, Berlin, New York: de Gruyter
- BUBENHOFER, NOAH, JOACHIM SCHARLOTH (2015), Maschinelle Textanalyse im Zeichen von Big Data und Data-driven Turn – Überblick und Desiderate, in: Zeitschrift für Germanistische Linguistik 43/1, 1–26
- CHEN, CHUN-HOUH, WOLFGANG HÄRDLE, ANTONY UNWIN (Hg.) (2008), Handbook of data visualization, Berlin: Springer
- COLEMAN, E. GABRIELLA (2012), Coding Freedom: The Ethics and Aesthetics of Hacking, Princeton: Princeton University Press
- EVERT, STEFAN (2009), 58. Corpora and collocations, in: LÜDELING, ANKE, MERJA KYTÖ (Hg.), Corpus Linguistics. An international handbook, Bd. 2, Berlin, New York: Mouton de Gruyter, 1212–1248
- FELDER, EKKEHARD, MARCUS MÜLLER, FRIEDEMANN VOGEL (2011), Korpuspragmatik: Thematische Korpora als Basis diskurslinguistischer Analysen, De Gruyter.
- FLECK, LUDWIG (1983), Erfahrung und Tatsache: gesammelte Aufsätze, hg. von LOTHAR SCHÄFER und THOMAS SCHNELLE, Frankfurt am Main: Suhrkamp
- FORD, PAUL (2015), What Is Code? If You Don't Know, You Need to Read This, in: Businessweek June 11, <http://www.bloomberg.com/whatiscode/> (zugegriffen am 8.7.2015)
- GEERTZ, CLIFFORD (1987), Dichte Beschreibung. Bemerkungen zu einer deutenden Theorie von Kultur, in: Dichte Beschreibung. Beiträge zum Verstehen kultureller Systeme, Frankfurt am Main: Suhrkamp, 7–43
- GEOFFEY, ANDREW (2014), Technology, Logistics and Logic: Rethinking the Problem of Fun in Software, in: GORIUNOVA, OLGA (Hg.), Fun and software: exploring pleasure, paradox, and pain in computing, New York: Bloomsbury Academic, 21–40
- GÖRG, CARSTEN, ZHICHENG LIU, JOHN STASKO (2014), Reflections on the evolution of the Jigsaw visual analytics system, in: Information Visualization 13/4, 336–345
- KEIBEL, HOLGER, CYRIL BELICA (2007), CCDB: A Corpus-Linguistic Research and Development Workbench, <http://corpora.ids-mannheim.de/cl2007-134.pdf>
- KEIM, D. A. u. a. (2006), Challenges in Visual Data Analysis, <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1648235>
- KEIM, D. A. u. a. (2010), Mastering the Information Age – Solving Problems with Visual Analytics, Goslar: Eurographics Association
- KRÄMER, SYBILLE (2009), Operative Bildlichkeit. Von der ›Grammatologie‹ zu einer ›Diagrammatologie‹?, in: HESSLER, MARTINA, DIETER MERSCH (Hg.), Logik des Bildlichen. Zur Kritik der ikonischen Vernunft, Bielefeld: Transcript, 94–123
- MANOVICH, LEV (2014), Software is the Message, in: Journal of Visual Culture 13/1, 79–81
- MAUTNER, GERLINDE (2012), Die kritische Masse. Korpuslinguistik und kritische Diskursanalyse, in: FELDER, EKKEHARD u. a. (Hg.), Korpuspragmatik: thematische Korpora als Basis diskurslinguistischer Analysen, Berlin: De Gruyter
- MIKOLOV, TOMAS u. a. (2013), Efficient Estimation of Word Representations in Vector Space, in: arXiv:1301.3781 [cs], <http://arxiv.org/abs/1301.3781> (zugegriffen am 14.4.2016)
- MORETTI, FRANCO (2000), Conjectures on World Literature, in: New Left Review 1, 54–68
- MORETTI, FRANCO (2009), Kurven, Karten, Stammbäume. Abstrakte Modelle für die Literaturgeschichte, Frankfurt am Main: Suhrkamp
- PEIRCE, CHARLES S. (1994), The collected papers of Charles Sanders Peirce, hg. von CHARLES HARTSHORNE u. a., Charlottesville, Va.: InteLex Corp., <http://pastmasters2000.nlx.com> (zugegriffen am 14.1.2016).
- PFEFFER, JÜRGEN (2010), Visualisierung sozialer Netzwerke, in: STEGBAUER, CHRISTIAN (Hg.), Netzwerkanalyse und Netzwerktheorie, Wiesbaden: VS Verlag für Sozialwissenschaften, 227–238
- RIEDER, BERNHARD, THEO RÖHLE (2012), Digital Methods: Five Challenges, in: BERRY, DAVID M. (Hg.), Understanding Digital Humanities, Basingstoke: Palgrave Macmillan, 67–84

- SCHARLOTH, JOACHIM, DAVID EUGSTER, NOAH BUBENHOFER (2013), Das Wuchern der Rhizome. Linguistische Diskursanalyse und Data-driven Turn, in: BUSSE, DIETRICH, WOLFGANG TEUBERT (Hg.), Linguistische Diskursanalyse. Neue Perspektiven, Wiesbaden: Springer VS, 345–380
- SHNEIDERMAN, BEN (1996), The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations, http://www.interactiondesign.us/courses/2011_AD690/PDFs/Shneiderman_1996.pdf
- SPITZMÜLLER, JÜRGEN, INGO H. WARNKE (2011), Diskurslinguistik: eine Einführung in Theorien und Methoden der transtextuellen Sprachanalyse, Berlin: de Gruyter
- THOMAS, JAMES J., KRISTIN A. COOK (Hg.) (2005), Illuminating the Path: The Research and Development Agenda for Visual Analytics, Los Alamitos, Calif.: National Visualization and Analytics Ctr.